Секция «Математические модели в космосе»

Сравнительный анализ эффективности различных моделей в задаче нахождения голосовой активности

Научный руководитель – Шишкин Алексей Геннадиевич

Литвинов Иван Борисович
Студент (специалист)
Московский государственный университет имени М.В.Ломоносова, Факультет космических исследований, Москва, Россия
E-mail: litvan007@gmail.com

В настоящее время, когда устройства, управляемые с помощью голоса, используются чрезвычайно широко, в том числе и на орбитальных станциях, распознавание голосовой активности приобретает особую важность. В данной работе рассмотрена задача детектирования голосовой активности в аудиосигнале в условиях высокой зашумленности, что является характерным при управлении механическим манипулятором на международной космической станции (МКС) с помощью голоса. В качестве решения были разработаны две модели, в первой из которых использовалась глубокая сверточная нейронная сеть, а во второй применялся модифицированный классический алгоритм [1].

Первая модель [2], которая использовала признаки eGemapsv02 [3], выделенные с помощью библиотеки OpenSMILE [4], состояла из трех двумерных сверточных слоев, с последующим преобразованием данных в массив длины отсчетов сигнала с помощью трех транспонированных сверточных слоев. Вторая модель, основан ная на классическом подходе оценки частот, рассматривала комбинацию функции среднего числа переходов через ноль и энергии сигнала.

Результаты, полученные при тестировании двух моделей на большом наборе аудиозаписей при разных отношениях сигнал/шум, подтверждают более высокую точность (precision) и f1-меру у сверточной нейронной сети (точность: 0.972, полнота: 0.927, f1-мера: 0.947) по сравнению с классическим подходом (точность: 0.553, полнота: 0.942, f1-мера: 0.665). Полнота в классическом подходе оказалась выше, потому что этот алгоритм склонен к более широкому обобщению, в то время как сверточная нейронная сеть выдает более точные результаты на определенных типах объектов. Таким образом, можно сделать вывод, что применение глубоких сверточных нейронных сетей является эффективным подходом к решению задачи распознавания голосовой активности с низким отношением сигнал/шум. Это открывает новые возможности при использовании устройств с голосовым управлением в условиях высокой зашумленности.

Источники и литература

1) Литвинов И. Б., Сазонова С. В. Создание систем голосового управления для механического манипулятора // II молодежные чтения имени М.В.Келдыша – 2022. – С. 23–39.
2) Процеров С. Д., Шишкин А. Г. Сегментация зашумленных речевых сигналов // Искусственный интеллект и принятие решений – 2021. №1. – С. 75-85.
3) Florian Eyben, Klaus R. Scherer, Bjorn W. Schuller, JohanSundberg, Elisabeth Andr 'e, Carlos Busso, Laurence Y. Devillers, Julien Epps, Petri Laukka, Shrikanth S. Narayanan, and Khiet P. Truong, "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing,"IEEE Transactions on Affective Computing, vol. 7, no. 2, pp.190–202, 2016.
4) openSMILE 3.0 - https://www.audeering.com/research/opensmile/